



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Seeing through the double blind

Citation for published version:

Killin, L & Della Sala, S 2015, 'Seeing through the double blind', *The Psychologist*, vol. 28, no. 4, pp. 288-291. <<http://thepsychologist.bps.org.uk/volume-28/april-2015/seeing-through-double-blind>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

The Psychologist

Publisher Rights Statement:

© This is the accepted version of the following article: Killin, L., & Della Sala, S. (2015). Seeing through the double blind. *The Psychologist*, 28(4), 288-291, which has been published in final form at <http://thepsychologist.bps.org.uk/volume-28/april-2015/seeing-through-double-blind>.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Seeing through the double blind

Lewis Killin^{1,2} and Sergio Della Sala^{1,3}

1. Human Cognitive Neuroscience, Department of Psychology, University of Edinburgh
 2. Alzheimer Scotland Dementia Research Centre, Department of Psychology, University of Edinburgh
 3. Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh
-

Brain Gym® is a controversial, exercise-based intervention designed to improve academic performance in children (Ritchie, Chudler & Della Sala, 2012). On the Brain Gym® web site, there is a section that states that they do not use randomised controlled trials (RCTs) to check its efficacy because they “question the ethics of not offering equal opportunities to all participants” (Brain Gym® International, 2011). However: regardless of how confident we are in the treatment – penicillin or Brain Gym® - RCTs are an inescapable means to figure out whether an intervention is efficacious or not. Furthermore, the main tenet behind RCTs is that they should be double-blind, meaning that neither the participants nor the experimenters know who is assuming the real compound and who the placebo.

This is certainly the case for drugs supposed to alleviate the symptoms of dementia. These drugs stand on years of development, research, phases of trials, assessment and analysis. The final phase of this research is based on the outcome of cognitive tests. If a treatment proves effective, it finds a place on chemistry shelves.

Recently, research revealed that drinking chocolate could be added to the list of treatments for dementia. Chocolate is a good source of flavonoids, naturally occurring compounds which have been associated with a staggering list of health benefits, from reducing the risk of cancer to increasing neuronal strength and connectivity.

In 2012, Desideri and colleagues investigated the effect of flavanol – as administered through “dairy-based cocoa drinks” – on cognitive function in patients at risk of dementia using a double-blind RCT. The result was a positive, dose-dependent effect of flavanols on some measures of executive function, all of which was packaged to represent positive changes in cognition.

The finding naturally found its way into the popular press, which was keen to report the excellent news that chocolate was scientifically proven to “halt dementia” (Willey, 2012). It was especially excellent news for Mars Inc., who funded the study and provided the dairy-based cocoa drinks. By capturing a beneficial effect of its product, this company could capitalise on the dementia epidemic.

[Figure 1]

The hunt for significant and positive outcomes for a product draws out the insidious bias of industry funding on study outcome. Specifically, a study that is funded by the same industry that creates the product under assessment is more likely to report positive and larger effect outcomes in its favour than studies carried without such funding.

In perspective, this bias might not be a severe cause for concern in the case of drinking chocolate. However, what if the same position is taken by an industry that could have serious implications for treatment or economical consequences for the health service?

We revealed this pattern of results for studies assessing donepezil, which is routinely prescribed as treatment for mild to moderate cases of Alzheimer’s Disease (AD). Here, industry-funded, double-

blinded RCTs that reported the effect of donepezil on cognitive measures, such as the Mini Mental State Examination (MMSE) or Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog), revealed, on average, a larger effect of the drug than the independent studies did (Killin, Russ, Starr, Abrahams & Della Sala, 2014). This difference remained after controlling for different study lengths.

[Figure 2]

Our finding is not a one-off. The bias of funding is the ubiquitous bad penny of pharmaceutical trials and intervention studies, and has been for some time. A 2003 review commented that an estimated 23-28% of biomedical researchers were thought to receive funding from the industry, and 34% of studies published in 1993 in major medical journals had such researchers as lead authors (Bekelman, Li & Gross, 2003). The consequence of this is apparent: Lundh, Sismondo, Lexchin, Busuioc and Bero (2012) reported that clinical trials funded by pharmaceutical companies were more likely to reveal a positive outcome for treatment over placebo than independent trials. In nutritional trials, health effects of non-alcoholic drinks were successfully revealed at a rate of 63% in non-industry intervention studies, which pales in contrast to the 100% success rate purported by industry-funded research (Lesser, Ebbeling, Gozner, Wypil & Ludwig, 2007).

The worst of this bias comes when it is not separated or acknowledged in bodies of evidence like meta-analyses or systematic reviews. In this case, a quantitative meta-analysis collates all observations of a product's effect under some particular situation – like double-blinded RCTs – to give an average indication of its effect over a placebo or control, independently of funding sources.

When independent and industry-funded studies are amalgamated into a single effect the combined effect is, at best, slightly altered; at worst, entirely misleading. Cataldo, Prochaska and Glantz (2010) teased apart the significant risk factor of tobacco on AD reported by independent studies away from the null – even *protective* – effects reported by the tobacco industry's research. The contrasting stories that these data provide meant that “if one simply combined all 43 studies in a single random effects meta-analysis, one would obtain an inaccurate null result,” (Cataldo *et al.*, 2010, p. 5), implying that tobacco smoke was not a risk factor for AD.

[Figure 3]

The proof and prevalence of this bias inspires scepticism for any data pushed by a group that could profit from convincing you their product works. This is not restricted to hard health data. Psychology studies house evidence for a plethora of educational or generic cognitive training interventions, and the chance of funding bias still looms.

A recent case in point is Cogmed, which is a brain-training programme that claims to exploit the plasticity of neural working memory systems to improve learning difficulties associated with Attention Deficit Hyperactivity Disorder (ADHD). Recently, it has drawn particular ire from parents and unimpressed psychologists (Colby, 2014). These sentiments echoed a lacklustre review that concluded it may, at best, be “possibly efficacious” given that it was often contrasted with poor or inadequate control groups (Chacko *et al.*, 2013a). Further criticism came from evidence of its limited effect on general working memory abilities and failure to improve parent-reported ADHD symptoms (Chacko *et al.*, 2013b). These claims, however, contrast clearly with earlier studies funded by and written by authors with shares in Cogmed (Klingberg *et al.*, 2005; Klingberg, 2012).

Back to Brain Gym®: it is worth stressing that the positive data reported may represent vested interests. As laid out plainly by Spaulding, Mostert and Beam (2010), the majority of the evidence detailing Brain Gym®'s efficacy was published in either the *Brain Gym® Journal* or the *Brain Gym® Magazine*. When assessed formally by means of a systematic review, there was no evidence from peer-reviewed

journals to suggest Brain Gym® worked, or that these studies had employed sound methods (Hyatt, 2007; Ritchie et al., 2012).

These shortcomings are not specific to Cogmed or Brain Gym®, but rather serve as examples of the general attitude towards the brain training industry. As raised by Owen *et al.* (2010) and Thompson *et al.* (2013), brain training produces improvement on the training exercise used in the experiment, but does not result in any statistically significant transfer to other skills.

Cynically, we may not be so surprised. Scientifically, though, we should be puzzled. How does this bias persevere even after studies are put through the wringer of double-blinded RCTs? The logic of such trials is that most confounding variables are taken out in the wash, giving clean results. Where does the industry's stream of good news come from?

Publication and design bias

Plainly, this bias may simply be another flavour of publication bias. The prevalence of publication bias in cognitive sciences has been carefully detailed and reviewed by Ioannidis, Munafò, Fusar-Poli, Nosek and David (2014), who noted that, in spite of a near-even split between negative and positive industry-funded trials investigating the effect of antidepressants, the former are suppressed and unpublished (Turner *et al.*, 2008). In the case that these negative findings *might* be published, however, we would expect these studies to surface in journals well after positive findings (Misakian and Bero, 1998), affecting the rate at which we update our knowledge.

However: this presupposes that all trials are seen through to the end. It may be more cost-effective to call a poor project to a halt if the initial results are shaky. In this way, publication bias may not explain everything, but *design* bias may.

Fries and Krishnan (2004), observed in their assessment of American College of Rheumatology meeting abstracts that 100% of the total number of abstracts for RCTs of the study's sponsor's drug resulted in a positive outcome. In their discussion, they argue that the ceiling effect seen here could not be attributable to publication bias alone. Instead, they posit that design bias may be the most powerful explanation.

Specifically, as a drug is screened through multiple phases of testing, its weaknesses and strengths are revealed to researchers within the industry. Trials could then be designed to both capitalise on strengths and shade weaknesses. Doing so is a violation of the principle of equipoise, where a drug's efficacy – in comparison to a control – should always be uncertain, and therefore necessitate research. Djulbegovic *et al.* (2000) revealed that the pharmaceutical industry violates this principle where independent organisations do not. Coupled with the study's finding of the industry's preference to test a product against a placebo or no therapy (as opposed to an active or competing treatment) this suggests that industry studies are designed to be "safe".

Variable of interest

Psychologists need to pay particular attention to the scale or test they use to demonstrate efficacy. This is certainly the case in drug trial literature, as multiple scales can be used to assess one aspect of disease; it would be wrong to assume that these are all equivalent. Indeed, this was the case in trials on donepezil; cognition was assessed by both the MMSE and the ADAS-cog, yet without any clear rationale behind each choice. On the one hand, each scale has its own strengths and weaknesses, but on the other they broadly assess the same construct; does the choice of scale really have a bearing on the outcome?

An analysis of available data suggests that it does. We compared ADAS-Cog and MMSE as outcome measures using standardized mean difference (SMD) which allows one to assess outcome when measured with different tools. After separating trials based on their outcome measure, we see that the ADAS-cog is associated with a much greater SMD compared to placebo than the SMD associated with

the MMSE, meaning that ADAS-Cog shows greater effects of the drug than the MMSE. Both scales ostensibly measure cognition, but clearly not to the same effect. Moreover: which scale should we trust?

This difference may be explained by the fact the ADAS-cog has been criticized for being too easy (Cressey, 2012). If this is the case, donepezil participants assessed on this scale may always find themselves close to ceiling level despite exhibiting subtle declines that a different scale would reveal. In other words: donepezil groups may still be declining – just not at a rate the ADAS-cog is designed to detect.

Thus, two scales of cognition, often used interchangeably, and considered equipoise in reviews and meta-analyses, do not necessarily present the same effect. It is noticeable that the drinking chocolate study from Desideri *et al.* (2012) observed a change in speed in executive tasks, but no difference in the MMSE (which, crucially, assesses neither executive function nor speed). The change in the former is still reported as “cognitive function”. Strictly, this is true, but it is disingenuous to use a broad term that could be misinterpreted so as to apply to a wider range of skills. Ultimately, our broad definitions give ample room for researchers to use and choose tasks that are falsely portrayed as tantamount, yet produce different results. The effect of interventions will vary accordingly. Psychologists should be aware of these potential biases.

Data creativity

In spite of careful planning, there is no guarantee that a trial will work. A failed experiment could be an expensive failure, both in terms of time and money. What is your resolve if a three-year, multi-centre trial did not pan out?

If there are enough collected data or possible outcome variables, there is a good opportunity to find or massage the desired effect in spite of what is declared *a priori*. Antidepressant trials reviewed by Turner and colleagues (2008) demonstrated that some industry-funded studies failed to find a statistically significant effect on the primary outcome they had specified in the methods they submitted to the US Food and Drug Administration. Instead, a different result was promoted as the significant, primary finding, and the original variable of interest was shelved.

Qualitatively, this has been examined in an unsettling historical case analysis from Cataldo, Bero and Malone (2010), which investigated correspondence between the tobacco industry and investigators of the Framingham Heart Study. The introduction of the Council for Tobacco Research – and its money – to the study came at the cost of the Framingham data, which would later be reinterpreted and reanalysed to conclude that there was no effect of tobacco smoke on coronary heart disease (CHD). Specifically, an analysis of these data stratified by age and ethnicity would suggest some individuals were naturally predisposed to CHD, and divert attention away from the hypothesis that tobacco was a cause. The latter position was pushed by the original investigators and supported by reference to overall mortality data. In the end, the same data produced two contrasting conclusions.

Ignoring criticism

Positive effects produced by the industry may be due to significant methodological shortcomings in spite of using the RCT method. In these cases, reviewers and critics need to identify these issues. The industry, however, has to listen.

A failure to do so became clear when the health benefits of transcendental meditation (TM) programmes were held under scrutiny. Its use pointed towards possible positive effects on mental health and blood pressure. However, a series of meta-analyses reviewed these claims in turn. Firstly, Canter and Ernst (2003) observed a clear divide between studies that claimed a beneficial effect of TM on cognition and those that did not. Specifically, TM’s success was linked to inadequate control groups (echoing the findings of Chacko *et al.*, 2014 and Djulbegovic *et al.*, 2000). On TM’s link with blood pressure, Canter and Ernst (2004) argued that, in a field saturated by TM affiliated studies, even the most stringent RCT

failed to adequately assess baseline measures and also account for differences between experimental groups, such as rates of medication. This was raised as a serious criticism and a source of bias in the TM literature.

After these criticisms were published, Anderson, Liu and Kryscio (2008) produced an updated meta-analysis of TM and blood pressure. This naturally contained the criticised studies analysed by Canter and Ernst four years previously. The authors plainly concluded that the quality of TM studies varied. However, given this range of quality their approach was to highlight the high quality studies, and ignore the fundamental problems raised by Canter and Ernst. The authors nonetheless concluded that TM was responsible for clinically meaningful changes in systolic and diastolic blood pressure. Unlike previous meta-analyses, this was funded by an unlimited gift from Dr. Howard Settle, who is responsible for personally funding the construction of TM centres in North America.

Psychology's role

So: how do we get to the truth of a product's effectiveness? We may be starting on poor footing as it is. Psychology is fraught with biases and possible confounding effects. For instance, even well-established phenomena in short term memory - such as the phonological similarity or word length effects - are affected by the strategies participants use (Logie *et al.* 1996). How many other possible variables could explain a part of a human's cognition or behaviour? What if participants are anxious or if they are not familiar with the words we ask them to remember? At what point do we stop controlling for these variables?

We cannot - and do not - address all confounds. Instead, we control for the bigger, broader variables like age, education or gender and work on the assumption that multiple measurements merge and overlap to reveal some identifiable construct or mechanism that can be generalised (Lilienfield, 2012), and rely on replications (as emphasised in *The Psychologist*, May 2012) to make sure our finding is not restricted to one place and time (as Manzi, 2010, would argue). The least we can do then, in addition to this, is to control for the bias of industry funding, especially if it is something that "researchers cannot stop" (Seife, 2012, p. 63).

Such control has started to take place in Europe. This year, MEPs voted in favour of the Clinical Trials Regulation, which was created to ensure that trials are pre-registered before they are run, and that the results are clearly reported and explained in lay terms within a year of the study's end. This takes away the likelihood of publication bias, design bias and *post hoc* massaging as detailed earlier. Pre-registration may prove an effective tool also in cognitive science (Chambers, 2013; see discussion in *The Psychologist*, July 2013).

However, the scientific community cannot coast on the new regulation alone. Researchers will still need to be relentlessly critical of industry-funded research. The ammunition closest to hand for a psychologist would be their knowledge of statistics and experimental design, but an education in psychometrics can define the overlap between an affected construct and the test used to assess it, cross-examining the exact claim of industry-funded studies.

Psychologists are also in a position to take on the industry with their own research.

This year marks the 50th anniversary of the 1964 *Surgeon General's Report on Smoking and Health*. To commemorate this event, the role of psychologists in the continued campaign for tobacco control was reviewed recently (De Angelis, 2014). Amongst other contributions, psychologists were responsible for developing population-wide smoking cessation interventions, such as telephone quitting lines, which are readily available as government funded amenities across the United States. Furthermore, related psychological interventions have been carefully developed to be sensitive to different populations so as to identify the exact relationships between groups and outcomes associated with their smoking. These

efforts stand in contrast to the evidence surrounding nicotine patches, which is a generic treatment with an efficacy affected by industry-funding (Etter, Burri & Stapleton, 2007).

Conclusions

Funding sources will continue to bias scientific research. The move towards trial registration will take the spin off industry-funded research, but there is still a need for diligence. As this bias perseveres, researchers need to remain critical and forge a way around this bias or, at least, flag it at every instance.

References

- Anderson, J. W., Liu, C., & Kryscio, R. J. (2008). Blood pressure response to transcendental meditation: a meta-analysis. *American Journal of Hypertension*, 21(3), 310-316.
- Bekelman, J. E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA*, 289: 454-465.
- Cataldo, J. K., Prochaska, J. J., & Glantz, S. A. (2010). Cigarette smoking is a risk factor for Alzheimer's Disease: an analysis controlling for tobacco industry affiliation. *Journal of Alzheimer's Disease*, 19(2), 465-480.
- Cataldo, J. K., Bero, L. A., & Malone, R. E. (2010). "A delicate diplomatic situation": tobacco industry efforts to gain control of the Framingham Study. *Journal of Clinical Epidemiology*, 63(8), 841-853.
- Canter, P. H., & Ernst, E. (2004). Insufficient evidence to conclude whether or not Transcendental Meditation decreases blood pressure: results of a systematic review of randomized clinical trials. *Journal of Hypertension*, 22(11), 2049-2054.
- Canter, P. H., & Ernst, E. (2003). The cumulative effects of Transcendental Meditation on cognitive function—a systematic review of randomised controlled trials. *Wiener Klinische Wochenschrift*, 115(21-22), 758-766.
- Chacko, A., Bedard, A. C., Marks, D. J., Feirsén, N., Uderman, J. Z., Chimiklis, A., ... & Ramon, M. (2013). A randomized clinical trial of Cogmed Working Memory Training in school-age children with ADHD: a replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*, 55(3), 247-255.
- Chacko, A., Feirsén, N., Bedard, A. C., Marks, D., Uderman, J. Z., & Chimiklis, A. (2013). Cogmed working memory training for youth with ADHD: A closer examination of efficacy utilizing evidence-based criteria. *Journal of Clinical Child & Adolescent Psychology*, 42(6), 769-783.
- Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49(3), 609-610.
- Colby, L. (2014, May 21). Brain Training Seen as Dud for Attention-Deficit Children. *Bloomberg*. Retrieved from <http://www.bloomberg.com/news/2014-05-21/brain-training-seen-as-dud-for-attention-deficit-children.html>
- Cressey, D. (2012, December 18). Alzheimer's test may undermine drug trials. *Nature News*. Retrieved from <http://www.nature.com/news/alzheimer-s-test-may-undermine-drug-trials-1.12079>
- DeAngelis, T. (2014). Thank you for not smoking. *Monitor on Psychology*, 45(3), 40.
- Desideri, G., Kwik-Urbe, C., Grassi, D., Necozione, S., Ghiadoni, L., Mastroiacovo, D., ... & Ferri, C. (2012). Benefits in Cognitive Function, Blood Pressure, and Insulin Resistance Through Cocoa Flavanol Consumption in Elderly Subjects With Mild Cognitive Impairment The Cocoa, Cognition, and Aging (CoCoA) Study. *Hypertension*, 60(3), 794-801.

- Djulbegovic, B., Lacey, M., Cantor, A., Fields, K. K., Bennett, C. L., Adams, J. R., ... & Lyman, G. H. (2000). The uncertainty principle and industry-sponsored research. *The Lancet*, 356(9230), 635-638.
- Etter, J. F., Burri, M., & Stapleton, J. (2007). The impact of pharmaceutical company funding on results of randomized trials of nicotine replacement therapy for smoking cessation: a meta-analysis. *Addiction*, 102(5), 815-822.
- Fries, J. F., & Krishnan, E. (2004). Equipoise, design bias, and randomized controlled trials: the elusive ethics of new drug development. *Arthritis Res Ther*, 6(3), R250-5.
- Hyatt, K. J. (2007). Brain Gym® Building Stronger Brains or Wishful Thinking? *Remedial and Special Education*, 28(2), 117-124.
- Ioannidis, J., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., & David, S. P. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in Cognitive Sciences*, 18(5), 235-241.
- Killin, L. O., Russ, T. C., Starr, J. M., Abrahams, S., & Della Sala, S. (2014). The effect of funding sources on donepezil randomised controlled trial outcome: a meta-analysis. *BMJ Open*, 4(4), e004083.
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... & Westerberg, H. (2005). Computerized training of working memory in children with ADHD-a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 44(2), 177-186.
- Klingberg, T. (2012). Is working memory capacity fixed? *Journal of Applied Research in Memory and Cognition*, 1(3), 194-196.
- Lesser, L. I., Ebbeling, C. B., Goozner, M., Wypij, D., & Ludwig, D. S. (2007). Relationship between funding source and conclusion among nutrition-related scientific articles. *PLoS Medicine*, 4(1), e5.
- Logie, R. H., Della Sala, S., Laiacina, M., Chalmers, P., Wynn, V. (1996) Group aggregates and individual reliability: The case of verbal short-term memory. *Memory and Cognition*, 24 (3): 305-321.
- Lilienfeld, S. O. (2012). Public skepticism of psychology: why many people perceive the study of human behavior as unscientific. *American Psychologist*, 67(2), 111-129.
- Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O. A., & Bero, L. (2012). Industry sponsorship and research outcome. *Cochrane Database Syst Rev*, 12.
- Manzi, J. (2012, Summer). What Social Science Does – and Doesn't – Know. *City Journal*. Retrieved from http://www.city-journal.org/2010/20_3_social-science.html
- Misakian, A. L., & Bero, L. A. (1998). Publication bias and research on passive smoking: comparison of published and unpublished studies. *JAMA*, 280(3), 250-253.
- Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., ... & Ballard, C. G. (2010). Putting brain training to the test. *Nature*, 465(7299), 775-778.

- Ritchie, S.J., Chudler, E.H. & Della Sala, S. (2012) Don't try this at school: the attraction of 'alternative' educational techniques. In Della Sala, S. & Anderson, M. (Eds.) *Neuroscience in Education: the good, the bad and the ugly*. Oxford University Press, 244-264.
- Seife, C. (2012). Is Drug Research Trustworthy? *Scientific American*, 307(6), 56-63.
- Spaulding, L. S., Mostert, M. P., & Beam, A. P. (2010). Is Brain Gym® an effective educational intervention? *Exceptionality*, 18(1), 18-30.
- Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... & Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS One*, 8(5), e63614.
- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358(3), 252-260.
- Wiley, J. (2012, August 14). Chocolate can halt dementia. *Daily Express*. Retrieved from <http://www.express.co.uk/news/uk/339655/Chocolate-can-halt-dementia>